FIRM/C781

LIBRARY
OCT 23 1985
National Marine Fisheries Service
2570 Dole Street
Honolulu, Hawaii

A MODIFICATION OF BHATTACHARYA'S METHOD FOR THE ANALYSIS OF MIXTURES OF NORMAL DISTRIBUTIONS

by

D. Pauly
International Center for Living Aquatic Resources Management (ICLARM)
MCC P.O. Box 1501
Makati, Metro Manila
The Philippines

and

J.F. Caddy
Marine Resources Service
Fishery Resources and Environment Division

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
ROME, March 1985

PREPARATION OF THIS DOCUMENT

It is frequently the case that a minor improvement to a previously existing method arises in the course of some practical application of existing theory that offers significant possibilities for facilitating a previously tedious or expensive procedure. While such improvements cannot be regarded as original contributions to science, they need to be recorded for the benefit of other users. It is in this context that the present procedure is presented. Other similar approaches that appear to offer savings in time and effort to fisheries assessment workers could be considered for publication in FAO Fisheries circulars.

ABSTRACT

The original version of Bhattacharya's method for the separation of mixtures of normal distributions requires the drawing of a graph, from which certain values (angles, intercepts) are read which are used for the estimation of the mean, standard deviation and size of each component distribution. As originally presented, the method is tedious and somewhat subjective.

A version of this method is presented here which makes this graph superfluous and uses an objective criterion to identify the component distributions in a mixture. The method is implemented for use with a HP 67/97 programmable calculator. Program listing and an application example are provided.

Note: "This replaces an earlier version of this circular which has a misprint on p. 11".

W/B4627

For bibliographic purposes this document should be cited as follows:

Pauly, D. and J.F. Caddy, 1985 A modification of Bhattacharya's method for the analysis of mixtures of normal distributions. FAO Fish.Circ., (781):16 p.

FAO Fisheries Circular (FAO Fish.Circ.)

A vehicle for distribution of short or ephemeral notes, lists, etc., including provisional versions of documents to be issued later in other series.

1. INTRODUCTION

Fishery biologists very frequently have to separate mixtures of several normal frequency distributions into their constituent parts. Such a task arises, for example, when length-frequency distributions need to be broken down into separate groups by age, sex or spawning period in order to make inferences on growth and mortality, of fish, or when various groups of oocytes, or larval stages in the plankton need to be separated by time of origin or growth rate.

The methods most commonly used for such a task are:

- the graphical methods of Harding (1949), Cassie (1954), and Tanaka (1962) and,
- the maximum likelihood method of Hasselblad (1966) as implemented in the NORMSEP (Abramson, 1971) and ENORMSEP (Yong and Skillman, 1971) computer programmes.

Except for the ENORMSEP programme, all these methods require an input as to the expected number of distributions; also, the graphical methods are more tedious than the method presented by us.

Bhattacharya's (1967) method for separating mixtures of normal distributions was proposed by its author as a simple graphical method, slightly superior to those proposed previously. It has not been used much, however; this may be due to the original account of the method, which is more complex than need be. Some (non-exhaustive) suggestions of how this procedure may be used in fisheries population analysis are given in an Appendix (1) to this document.

The version presented here of Bhattacharya's method is simplified as much as possible; also it is implemented on a programmable calculator, and incorporates an objective criterion for the identification of the various normal distributions.

1.1 Bhattacharya's Method

The method proposed by Bhattacharya consists of plotting the natural logarithms of the ratios of successive frequencies ($\log_e[f_{i+1}/f_i]$) on the midpoints of the corresponding length classes (x_i) to identify those parts of the plot that can be described by a straight line of the form

$$F_i = a + bx_i \qquad \ldots 1)$$

where $F_i = \log_e(f_i+1/f_i)$, and where the slope (b) has a negative sign.

In an earlier (unpublished) FORTRAN programme, Caddy proposed that such lines could be defined using three points only, and that their straightness should be tested by the degree of (negative) correlation between the F_i and the x_i values. This suggestion is incorporated in the method presented here.

When a straight line has been identified which goes through or close to 3 successive points and whose parameters a and b have been estimated (using the linear regression technique), the mean (μ), standard deviation (δ) and number of items (N) in the component distributions so identified are estimated from

$$\mu = \frac{(0.5d) - (a/b)}{} \qquad \ldots 2)$$

$$\delta = \sqrt{1/-b - (d^2/12)} \qquad \ldots 3)$$

and

$$N = \frac{\delta}{d} \exp\left[\frac{\Sigma \log_e F_{1,2,3}}{3} + \frac{\Sigma(x_{1,2,3} - \mu)^2}{6\,\delta^2 + (d^2/12} + \frac{d^2}{24\delta^2} + \log_e 2\pi\right] \qquad \ldots 4)$$

where $F_{1,2,3}$ and $x_{1,2,3}$ are the F_i and x_i values defining the three points used for the correlation and regression analyses, and where d is the class interval.

As presented here, Bhattacharya's method ceases to be a "graphical" method since the presence or absence of a normally distributed component is not inferred from the appearance of the plot over a given range of x values, but from the value of the correlation coefficient (r) of the points in that range of x-values. This new criterion also allowed us to turn Bhattacharya's method into an algorithm so short that it could be incorporated into a programmable calculator.

1.2. Bhattacharya's Method as a Calculator Programme

The first of the calculator programmes (Appendix II) presented below, which are implemented for use with a HP 67/97, has three main routines:

1. a routine to compute successive values of F_i, with a single entry of data,
2. a linear regression and correlation routine,
3. a routine to estimate μ, δ and N values for each triplet of F_i and x_i pairs with an absolute correlation coefficient higher than the critical value entered.

To use these routines, simply follow the instructions given in the example. Each frequency needs to be entered only once, and the values of μ, δ and N will be computed automatically every time a component distribution has been identified. Bhattacharya's method has its problem and pitfalls; here are three of them:

a. Zero frequencies must not be entered (zero has no log!); rather, the data must be regrouped such that every class considered has a non-zero frequency. However, also consider that the method is approximate, and works well only when the component distributions do not overlap too much, so that for occasional (single) blanks in a size frequency, inserting a very small (non-zero) value (e.g., 1, .1, .001, etc.), while mathematically inexact, may be the only way to avoid doubling class intervals and consequent loss of valuable information.

b. A component distribution which overlaps little with its neighbours might have more than 3 points forming a straight line. The programme therefore might output a run of two or more sets of μ, δ and N values, all pertaining to the same component. If these differ slightly, choosing those outputs with the highest value of r, or even plotting the values of $\log_e (f_i+1/f_i)$ over that part of the distribution, are possible alternative strategies. Testing alternative values with the simulated distribution generated in the second programme is also a possibility.

c. The critical value of r used might be so high as to cause genuine distributions to be ignored (Type II error), or so low as to cause irrelevant bumps to be misidentified as genuine components (Type I error). It is therefore recommended to run the data at least twice, using different critical values of r.

1.3 Improving and Testing of Results

Even when considering the caveats mentioned above, the method might still fail to identify all components of a mixture, especially when they greatly overlap.

Additional components can be detected, however, by comparing the raw frequencies ($\hat{f}_i$) with the expected frequencies ($\hat{f}_i$) that is with the frequencies that are computed from the means and standard deviations of the component identified so far. The equation used to compute the expected frequencies is

$$\hat{f}_i = \sum_{1}^{j} \hat{f}_i = \sum_{1}^{j} \frac{N_j d}{\delta_j \sqrt{2\pi}} \exp.-0.5\left[x_i-\bar{x}_j/\delta_j\right]^2 \qquad \ldots 5)$$

where μ_j and δ_j are the parameters of the components (j) (of number Nj identified), (d is the class interval) and x_i is the class midpoint corresponding to f_i. The computation of expected values allows for the computation of residuals (R), defined as:

$$R = f_i - \hat{f}_i \qquad \ldots 6)$$

An example follows of the use of the two programmes in Appendix II and III to identify normal components in an artifical mixture generated using input values (Table 2) from Everett and Hand (1981).

The two programmes may be used in a flexible way to investigate any mixture of normal components using programme 1 to identify components and programme 2 to generate predicted size frequencies and chi square statistics from the output of programme 1.

Usually, only a second run of Programme 1 will be sufficient to identify any component that was missed in the first run (see example: Table 2).

Once all components of a mixture of distributions have been identified, there is the possibility, using the chi-square test, to assess whether the predicted frequencies, as a whole, match the raw frequencies. To perform this test, proceed as follows:

1. compute the expected frequencies for each value of x_i,
2. use the expected and raw frequencies to compute, for each value of x_i, the quantity $(f_i-\hat{f})^2/\hat{f}_i$
3. add up the quantities obtained in (2) for all values of x_i
4. compare the sum obtained in (3) with the table values of the chi-square statistics; the number of degrees of freedom (d.f.) to use is equal to the number of classes considered (n) minus one, minus the number of component distributions (j) used in generating the $\hat{f}_i$-values, i.e., d.f. = n-(1+2j)

This computation of the chi-square statistic is documented in Table 1; note however that classes with $f_i < 5$ were pooled, to prevent inflating the test statistics computed in (3), as suggested in most statistics texts.

REFERENCES

Abramson, N.J., 1971 Computer programmes for fish stock assessment. FAO Fish.Tech.Pap., (101):pag.var.

Bhattacharya, C.G., 1967 A simple method of resolution of a distribution into Gaussian components. Biometrics, 23:115-35

Cassie, R.M., 1954 Some use of probability paper for the graphical analysis of polymodal frequency distributions. Aust.J.Mar.Freshwat.Res., 5:513-22

Everitt, B.S. and D.J. Hand, 1981 Finite mixture distributions. London, Chapman and Hall, 143 p.

Harding, J.F., 1949 The use of probability paper for the graphical analysis of polymodal frequency distribution. J.Mar.Biol.Assoc., U.K. 28:141-53

Hasselblad, V., 1966 Estimation of parameters for a mixture of normal frequency distributions. Technometrics, 8:431-46

Sokal, R.R. and F.J. Rohlf, 1981 Biometry, 2nd ed., San Francisco, W.H. Freeman and Company

Tanaka, S., 1962 A method of analysing polymodal frequency distributions and its application to the length distribution of the porgy, Taius tumifrons (T. and S.). J.Fish. Res.Board Can., 19(6):1143-59

Yong, M.Y. and R.A. Skillman, 1975 A computer program for analysis of polymodal frequency distributions (ENORMSEP), FORTRAN IV. Fish.Bull.NOAA/NMFS, 73(3):681

Table 1

Analysis of a set of frequency data (f_i) using Bhattacharya's method (see text for interpretation)

original data		first run a/		2nd run b/	
x_i	f_i	$\hat{f}_i$	$f_i-\hat{f}_i$	$\hat{f}_i$	$f_i-\hat{f}_i$
0.5	2	.99	.01	.99	1.01
1.5	23	27.17	4.17	27.17	-4.17
2.5	22	20.37	1.63	20.37	1.63
3.5	1	.66	.65	.66	.65
4.5	37	33.08	3.92	33.08	3.92
5.5	28	33.09	5.09	33.09	-5.09
6.5	1	.79	.21	.79	.21
7.5	6	5.67	.33	5.67	.33
8.5	17	17.74	-.74	17.74	-.74
9.5	16	16.49	-.49	16.49	-.49
10.5	5	4.56	.44	4.56	.44
11.5	2	0.37	1.63	.38	1.63
12.5	3	0.01	2.99	.06	.06
13.5	6	0.00	6	.97	5.03
14.5	6	0.00	6	5.90	.10
15.5	15	0.00	15	11.91	3.09
16.5	6	0.00	6	8.02	-2.02
17.5	2	0.00	2	1.80	.20
18.5	2	0.06	2	.13	.03

Table 2

Results of the analysis of the frequencies in Table 3 using Bhattacharya's method

		values estimated			input values used to generate f_i's c/	
		μ	δ	N	μ	δ
1st run a/						
	1st group	1.92	0.527	49.3	2.0	0.5
	2nd group	5.00	0.452	69.1	5.0	0.5
	3rd group	8.94	0.908	45.4	9.0	1.0
2nd run b/						
	4th group	15.64	0.954	28.8	15	2.0

a/ using r = 0.997 (i.e., 95% level) b/ using r = 0.988 (i.e., 90% level)
c/ as given in Everett and Hand (1981)

Computation example based on raw frequencies (f_i) in Table 3.

Programme 1

Step 1: Enter, or read card for first programme.

Step 2: Enter lower and upper limit of smallest size class (which in our example has a midpoint of 0.5, see Table 3):
Key strokes: 0↑1 fa

Step 3: Store critical value of correlation coefficient (for 95% level)
Key strokes 0.997 STO 8

Step 4: Enter first four frequency values
Keystrokes: 2↑23↑22↑1 R/S
Output: 1.916 (μ); 0.527(δ); 49.286 (N); i.e., component 1 has been identified.

Step 5: Enter succeeding frequencies, and identify the first set of components (e.g., Table 2)

Step 6: At this point, two possible (complementary) steps are possible:

Step 6A: Repeat steps 1-5 with programme 1 with a different (lower) critical value of the correlation coefficient. This may identify new components (see e.g., the 4th component in run 2, Tables 1 and 2). Note: some of these new, more questionable components may be identical with those picked up by run 1, or more than 1 (sequential) estimate may be identified for the same component, in which case, that with the highest r-value may be selected by iteration with different values of r.

PROGRAMME 2:

6B: Enter or read card for this programme.

This programme may now be used to generate numbers at size for each component identified individually. If only 1 set of μ, S+N values are entered, or if all are entered (e.g., run 2, Table 1) the predicted size frequencies ($\hat{f}_i$) will be generated. These can be compared against the original data and residuals ($f_i - \hat{f}_i$) calculated, again using programme 2.

Step 7: The residuals may be used as input to programme 1, to look for missing components, if the cumulative sum of residuals suggests that one or more has been overlooked in the first run. If an error message results from this procedure, one suggestion is to subtract the largest negative residual, (e.g., - (-5.09) in run 2, Table 1) from all residuals, and use the resultant positive series of numbers as input to Program 1. (Note: values of μ, S found this way may be reliable, but not of course the number of the component, N: it will be necessary here to use the new μ,S values as input to program 2 with different trial values of N, to minimize residuals and thus find the best value of N by iteration for the new component.

Step 8: After all components have been identified (see Table 2), recompute new expected frequencies as in step 6B using all sets of μ, δ and N values.

Step 9: Compute residuals and values of $(f_i - \hat{f}_i)^2/\hat{f}_i$ for use in computation of the chi-square statistic. Do not omit to lump classes with values of $\hat{f}_i < 5$ with adjacent classes (see Table 3).

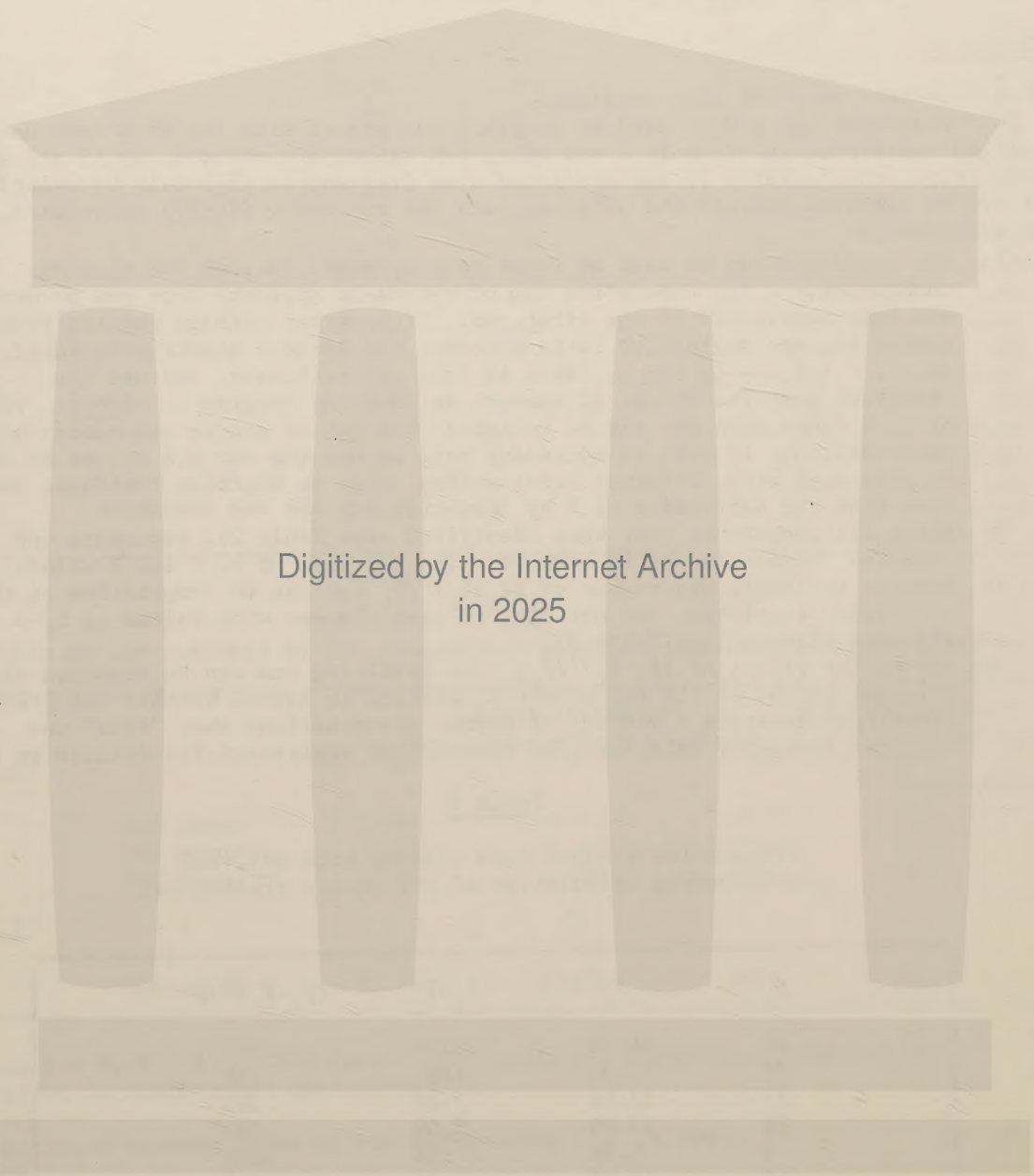
Step 10: Add up the values of $(f_i - \hat{f}_i)^2/\hat{f}_i$; the resulting sum can be compared with critical values of the chi-square statistics to assess whether the groups identified generate a mixture of normal distributions that "fits" the original frequency data (see any textbook of statistics for details on test.)

Table 3

Frequencies grouped into classes with $n \geq 5$, and corresponding calculation of chi square statistics.

f_i	$\hat{f}_i$	$f_i - \hat{f}_i$	$(f_i - \hat{f}_i)^2/\hat{f}_i$
25	28.16	3.16	.355
23	21.03	1.97	.185
37	33.08	3.92	.465
28	33.09	5.09	.783
7	6.46	.54	.045
17	17.74	.74	.031
16	16.49	0.49	.015
16	5.97	10.03	16.85
6	5.90	10	.002
15	11.91	3.09	.802
10	9.95	0.05	.001

Chi-square statistic = 19.54 for 11 classes and 2 d.f.'s



Digitized by the Internet Archive
in 2025

https://archive.org/details/fao-fisheries-and-aquaculture-circular_1985-03_781

APPENDIX 1

Some Applications of Bhattacharya's Method to Fish Population Analysis

The method as presently described, can of course be used to analyse single length frequencies to separate possible constituent components, and employed in this fashion can be useful in its own right, or as a precursor to analysis with a more 'rigorous' method, such as NORMSEP (Hasselblad, 1966). In this particular application, the means and variances obtained with the Bhattacharya method are used to provide initial guesses as to the number of components, and their means and standard deviations.

Concentrating however on the use of the HP program as a tool for population analysis, it might be helpful to provide some broad guidelines for use of the program, recognizing that:

(a) the main use of the program will be for population analysis; in particular, the analysis of growth and mortality of a population from a sequence of samples taken from a population over a period of time.

Sample size and class interval

(b) The way in which the samples are taken and measured will obviously have a significant impact on the types of inofrmation that can be extracted from them. In particular:

(1) class intervals should be small enough that the first few 'peaks' in the size distribution should be spanned by a minimum of three (preferably four or more) class intervals. This may not be possible to the right of the size distribution where due to slowing growth, the individual age groups show more overlap: which is a fundamental problem for all modal analysis methods.

(2) It is impossible to specify the number of individuals that need to be measured for all cases, but it should be related to the number of components that are likely to be separated: a figure of 100-200 per component is empirically correct, and this should also avoid embedded "blanks" in the distribution, which as already noted, poses problems with this method.

(3) If it is necessary to combine size frequencies from (e.g.) different types of gear, each with its own selectivity, fishing the same population; the weighing to be used should be in proportion to the estimated total weight captured by each type of gear.

Thus, if for class interval 1, the number of individuals measured from gears 1 and 2 are N_ℓ and $N_{\ell,2}$, the weights of samples are w_1 and w_2, and the estimated total catch weights in the time interval of interest are W_1 and W_2, the number in each class interval ℓ in the combined sample should be ideally given by:

$$N_{\ell,T} = (\frac{W}{w_1})N_{\ell,1} + (\frac{W2}{w_2})N_{\ell,2}$$

which is the 'expected' total number of individuals of that size caught by the fleet (all gears) in the interval, i.e., the "catch size frequency".

Growth rate estimation

Given a regular sampling throughout a full year by a gear that retains a significant number of (the youngest) age categories, the growth curve can be approximately reconstructed by "doubling up" the distribution (Figure 1), on the assumption that the size at age in a given season is roughly the same in successive years.

This allows at least a preliminary fitting of an overall growth curve to the data.

A certain degree of subjectivity is involved in fitting such a growth curve, in that the identity of age groups between successive time intervals (samples) needs to be maintained, in a situation where, due to sampling error, a given cohort may occasionally be missing, or a random 'bump' in the size frequency may be misidentified as corresponding to a cohort. One solution to this, is to run the Bhattacharya routine for 2 or 3 levels of the critical value of r, and give higher weighings to points that survive at higher critical r values, in drawing the growth curves as in Figure 1.

Mortality estimation

If the "catch size frequency" can be reconstructed as discussed earlier, and the identity of the same cohort maintained, then, an estimate of the overall mortality rate Z may be possible from the train of successive numbers (N_t, $N_{t+\Delta t}$, $N_{t+2\Delta t}$, etc.) for the same cohort. In theory, the corresponding annual rate of mortality $= \log_e \frac{(N_t - N_{t+\Delta t})}{\Delta t}$, where Δt is the sampling interval in fractions of a year. Given the fairly large errors that are likely to be involved in estimating (N_t), it would be better however to estimate Z from the slope of the log-linear regression of the numbers in a given component against time.

(Cautionary note: the method of "doubling up" the distribution described in Figure 1 is not to be used for mortality estimation).

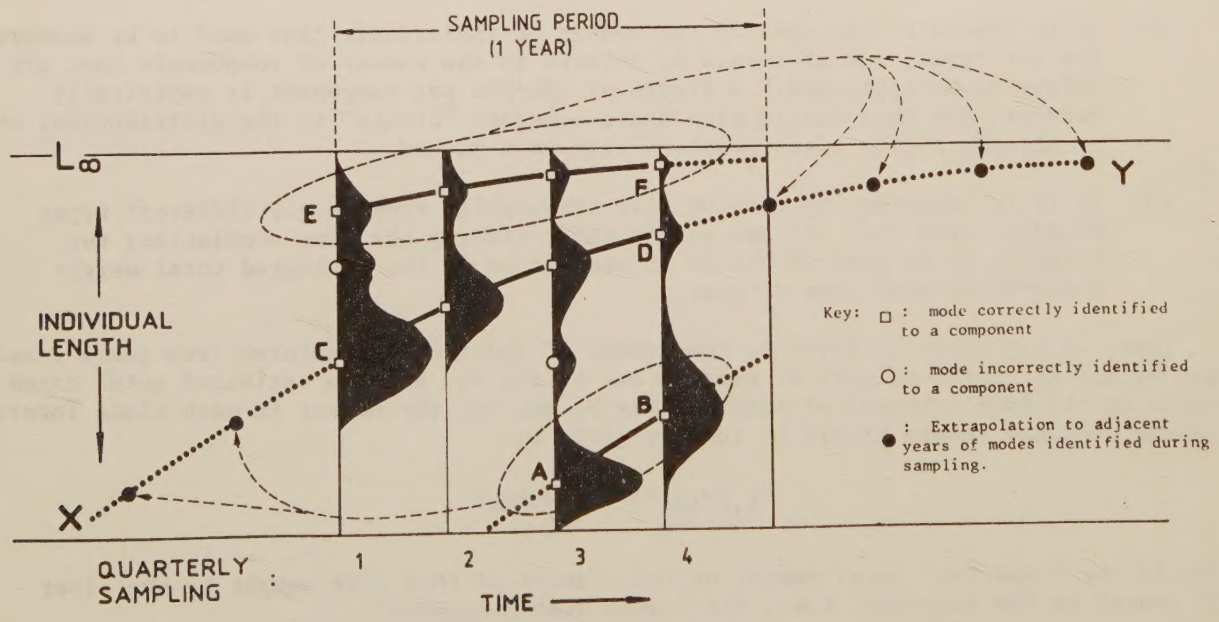
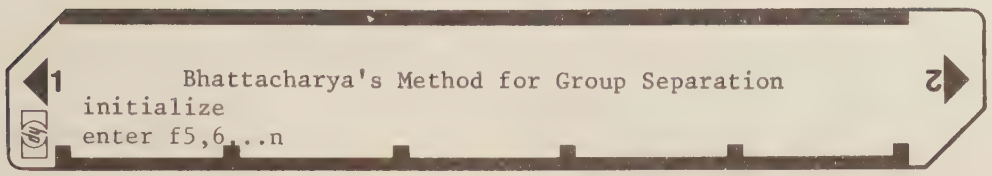


Fig. 1 Showing how segments of growth curves reconstructed from a single year of observations, may, with care, be "doubled up" to complete a growth curve over a larger time interval (see text).

User Instructions

APPENDIX II



STEP	INSTRUCTIONS	INPUT DATA/UNITS	KEYS	OUTPUT DATA/UNITS
1	Enter lower and upper limit of smallest size class considered	L_1 L_2	↑ f a	 0.000
2	Store critical value of correlation coefficient (1 d.f.) {See Program Description}	$r_{crit.}$	STO 8	$r_{crit.}$
3	Enter first four frequency values	f_1 f_2 f_3 f_4	↑ ↑ ↑ R/S	 0.000
	The display will either show zero, or the parameter values for the first group will be output, as follows:			or: μ (mean) δ (s.d.) N(numbers)
4	Enter fifth and all following frequencies	f_i	A	
	As in step (3), the output will be either zero or will consist of μ, δ and N for any group that has been identified.			

Note:

From the fifth value on, pressing RCL E will recall the last frequency value you have entered.

Program Description

Program Title Bhattacharya's method for group separation

Name Pauly, D. and J.F. Caddy
Address ICLARM, Manila and FAO, Rome
City **State** **Zip Code**

Date November 1982

Program Description, Equations, Variables, etc. This program is an application of Bhattacharya's method for the separation of normally distributed groups from a mixture of normal distributions. The method rests on the observation that when the natural logarithms ($\log_e$) of the ratio of successive frequencies (f_i) are plotted against class midpoints (x_i), a straight line with negative slope results when a group has been identified. Hence such a group can be defined by a line.

$$\log_e(f_{i+1}/f_i) = a + bx_i \qquad \ldots 1)$$

Here, the line is defined by 3 points only and its straightness is tested by the correlation between the $\log_e(f_{i+1}/f_i)$ and x_i value. When a straight line has been identified, the mean (μ), standard deviation (δ) and the number of items in the group in question are estimated using

$$\mu = (\frac{1}{2}d) - \frac{a}{b} \qquad \ldots 2)$$

$$\delta = \sqrt{1/-b-(d^2/12)}$$

and

$$N = \frac{\delta}{d} \exp \left\{ \frac{\Sigma \log_e(f_{i+1}/f_i)}{3} + \frac{\Sigma (x_{1,2,3,} - \mu)^2}{6\,\delta^2 + (d^2/12)} + \frac{d^2}{24\delta^2} + \log_e \sqrt{2\pi} \right\} \qquad \ldots 3)$$

where d is the class interval. The critical values of the correlation coefficient for use with this program are, with d.f. = 1, 0.997 (P = 0.05) and 0.988 (P = 0.10). (Lower test values - e.g., 0.9 - can be used if groups that are less well-defined are to be identified; this may lead to the "identification" of spurious groups, however).

Reference: Bhattacharya, C.G., 1967 Biometrics 23:115-35

Operating Limits and Warnings The method is approximate, and works well only when the groups that are to be separated do not overlap too much. When the groups overlap, it will be necessary to apply the method several times to a given data set, i.e., to subtract from the data the frequencies pertaining to groups that have been identified, then to apply the method again and so on. When "groups" are identified that have means very close to each other, it can be assumed that they pertain to the same group.

DO NOT USE THIS SPACE

Program Listing

			COMMENTS
001	001	*LBLo 21 16 11	
	002	CLRG 16-53	
	003	P⇄S 16-51	
	004	STOA 35 11	
	005	R↓ -31	
	006	STOB 35 12	
	007	RCLA 36 11	
	008	X⇄Y -41	
	009	- -45	
010	010	STO7 35 07	
	011	2 02	
	012	÷ -24	
	013	RCLB 36 12	
	014	+ -55	
	015	STO1 35 01	
	016	RCL7 36 07	
	017	+ -55	
	018	STO2 35 02	
	019	RCL7 36 07	
020	020	+ -55	
	021	STO3 35 03	
	022	CLX -51	
	023	R/S 51	
	024	STOD 35 14	
	025	R↓ -31	
	026	STOC 35 13	
	027	R↓ -31	
	028	STOB 35 12	
	029	R↓ -31	
030	030	STOA 35 11	
	031	RCLD 36 14	
	032	RCLC 36 13	
	033	÷ -24	
	034	LN 32	
	035	STO6 35 06	
	036	RCLC 36 13	
	037	RCLB 36 12	
	038	÷ -24	
	039	LN 32	
040	040	STO5 35 05	
	041	RCLB 36 12	
	042	RCLA 36 11	
	043	÷ -24	
	044	LN 32	
	045	STO4 35 04	
	046	GTO0 22 00	
	047	*LBLA 21 11	
	048	STOE 35 15	
	049	RCLD 36 14	
050	050	÷ -24	
	051	LN 32	
	052	STOI 35 46	
	053	RCLE 36 15	
	054	STOD 35 14	
	055	RCL5 36 05	
	056	STO4 35 04	
	057	RCL6 36 06	
	058	STO5 35 05	
	059	RCLI 36 46	
060	060	STO6 35 06	
	061	RCL7 36 07	
	062	ST+1 35-55 01	
	063	ST+2 35-55 02	
	064	ST+3 35-55 03	
	065	*LBL0 21 00	
	066	P⇄S 16-51	
	067	0 00	
	068	STO4 35 04	
	069	STO5 35 05	
070	070	STO6 35 06	
	071	STO7 35 07	
	072	STO8 35 08	
	073	STO9 35 09	
	074	P⇄S 16-51	
	075	RCL4 36 04	
	076	RCL1 36 01	
	077	Σ+ 56	
	078	RCL5 36 05	
	079	RCL2 36 02	
080	080	Σ+ 56	
	081	RCL6 36 06	
	082	RCL3 36 03	
	083	Σ+ 56	
	084	P⇄S 16-51	
	085	RCL8 36 08	
	086	RCL4 36 04	
	087	RCL6 36 06	
	088	x -35	
	089	RCL9 36 09	
090	090	÷ -24	
	091	- -45	
	092	ENT↑ -21	
	093	ENT↑ -21	
	094	RCL4 36 04	
	095	X^2 53	
	096	RCL9 36 09	
	097	÷ -24	
	098	RCL5 36 05	
	099	X⇄Y -41	
100	100	- -45	
	101	÷ -24	
	102	STOB 35 12	
	103	x -35	
	104	RCL6 36 06	
	105	X^2 53	
	106	RCL9 36 09	
	107	÷ -24	
	108	CHS -22	
	109	RCL7 36 07	
110	110	+ -55	
	111	÷ -24	
	112	$\sqrt{X}$ 54	

COMMENTS

REGISTERS

0 μ	1 L_1	2 L_2	3 L_3	4 Y_1/used	5 Y_2	6 Y_3	7 ΔL	8 r_{crit}	9 $d^2/12$
S0	S1	S2	S3	S4 Σx	S5 Σx^2	S6 Σy	S7 Σy^2	S8 Σxy	S9 3

A intercept/f_1	B slope/f_2	C $r_{amp.}/f_3$	D f_i	E f_{itl}	I δ

Program Listing

				COMMENTS
	113	STOC	35 13	
	114	RCL6	36 06	
	115	RCL4	36 04	
	116	RCLB	36 12	
	117	x	-35	
	118	−	-45	
	119	RCL9	36 09	
120	120	÷	-24	
	121	STOA	35 11	
	122	P≠S	16-51	
	123	RCLB	36 12	
	124	X<0?	16-45	
	125	GTO1	22 01	
	126	*LBL2	21 02	
	127	CLX	-51	
	128	RTN	24	
	129	*LBL1	21 01	
130	130	RCLC	36 13	
	131	RCL8	36 08	
	132	X>Y?	16-34	
	133	GTO2	22 02	
	134	RCL7	36 07	
	135	2	02	
	136	÷	-24	
	137	RCLA	36 11	
	138	RCLB	36 12	
	139	÷	-24	
140	140	−	-45	
	141	SPC	16-11	
	142	PRTX	-14	
	143	STO0	35 00	
	144	RCLB	36 12	
	145	CHS	-22	
	146	1/X	52	
	147	RCL7	36 07	
	148	X^2	53	
	149	1	01	
150	150	2	02	
	151	÷	-24	
	152	STO9	35 09	
	153	−	-45	
	154	$\sqrt{X}$	54	
	155	PRTX	-14	
	156	STOI	35 46	
	157	RCLD	36 14	
	158	TCL6	36 06	
	159	e^x	33	
160	160	÷	-24	
	161	STOC	35 13	
	162	RCL5	36 05	
	163	e^x	33	
	164	÷	-24	
	165	STOB	35 12	
	166	RCL4	36 04	
	167	e^x	33	
	168	÷	-24	

				COMMENTS
	169	STOA	35 11	
170	170	RCLB	36 12	
	171	x	-35	
	172	RCLC	36 13	
	173	x	-35	
	174	LN	32	
	175	3	03	
	176	÷	-24	
	177	STO4	35 04	
	178	RCL1	36 01	
	179	RCL0	36 00	
180	180	−	-45	
	181	X^2	53	
	182	RCL2	36 02	
	183	RCL0	36 00	
	184	−	-45	
	185	X^2	53	
	186	+	-55	
	187	RCL3	36 03	
	188	RCL0	36 00	
	189	−	-45	
190	190	X^2	53	
	191	+	-55	
	192	RCL9	36 09	
	193	RCLI	36 46	
	194	X^2	53	
	195	+	-55	
	196	6	06	
	197	x	-35	
	198	÷	-24	
	199	ST+4	35-55 04	
200	200	RCL9	36 09	
	201	RCLI	36 46	
	202	X^2	53	
	203	2	02	
	204	x	-35	
	205	÷	-24	
	206	ST+4	35-55 04	
	207	Pi	16-24	
	208	2	02	
	209	x	-35	
210	210	$\sqrt{X}$	54	
	211	LN	32	
	212	ST+4	35-55 04	
	213	RCL4	36 04	
	214	e^x	33	
	215	RCL7	36 07	
	216	÷	-24	
	217	RCLI	36 46	
	218	x	-35	
	219	PRTX	-14	
220	220	RTN	24	
	221	R/S	51	

LABELS

A $f_{5,6,\ldots,n}$ →	B	C	D	E
a initialize	b	c	d	e
0 used	1 used	2 used	3	4
5	6	7	8	9

FLAGS

0
1
2
3

SET STATUS

FLAGS	ON	OFF
0	☐	☒
1	☐	☒
2	☐	☒
3	☐	☒

TRIG	
DEG	☒
GRAD	☐
RAD	☐

DISP	
FIX	☒
SCI	☐
ENG	☐
n	3

Program Description

APPENDIX III

Program Title	Frequencies from a Mixture of Normal Distributions		
Name	Daniel Pauly	**Date**	May, 1983
Address	ICLARM, MCC P.O. Box 1501, Makati		
City	Metro Manila, The Philippines	State	Zip Code

Program Description, Equations, Variables, etc. This program uses the mean (μ_i), standard deviation (δ_i) and size (N_i) of a number of normal distributions ($1 \leq j \leq 6$) to estimate combined frequencies ($\hat{f}_i$).

The equation used is

$$\hat{f}_i = \sum_{1}^{j} \frac{N_j d}{\delta_j \sqrt{2\pi}} \exp.-0.5 \left| x_i - \bar{x}_j / \delta j \right|^2$$

where d is the class interval.

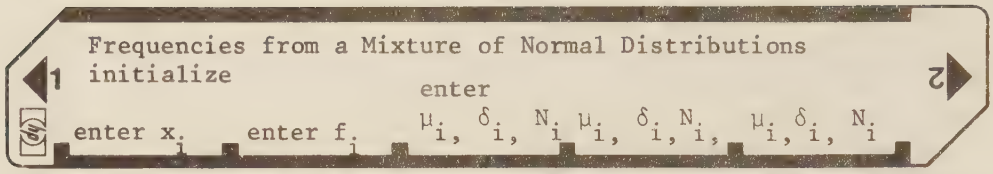
Residuals ($f_i - \hat{f}_i$) and values for the computation of the chi-square statistics $(f_i - \hat{f}_i)^2 / \hat{f}_i$ can also be computed.

Reference: Sokal, R.R. and F.J. Rohlf, Biometry. San Francisco, W.H. Freeman and Company (1981) 2nd Edition.

Operating Limits and Warnings When computing the chi-square statistics, do not use value of $\hat{f}_i$ smaller than 5; instead group several adjacent classes. It may be more convenient to hand-calculate values of $(f_i - \hat{f}_i)^2 / \hat{f}_i$ in this case. (See relevant section in textbook of statistics).

DO NOT USE THIS SPACE

User Instructions



STEP	INSTRUCTIONS	INPUT DATA/UNITS	KEYS	OUTPUT DATA/UNITS
1	Enter class interval and initialize	ΔX	f a	0.000
2	Enter sequentially the means (μ), standard deviations (δ) and size (N) of each component	μ_1	↑	
	(up to 6 sets of values can be entered).	δ_1	↑	
		N_1	C	1
		μ_2	↑	
	Note: do not enter anything beyond parameters of available components.	δ_2	↑	
		N_2	D	2
		μ_3	↑	
		δ_3	↑	
		N_3	E	3
		μ_4	↑	
		δ_4	↑	
		N_4	f C	4
		μ_5	↑	
		δ_5	↑	
		N_5	f d	5
		μ_6	↑	
		δ_6	↑	
		N_6	f e	6
3	Enter x_i value to estimate $\hat{f}_i$	x_i	A	$\hat{f}_i$
4	To calculate residual and values for the chi-square statistics, enter actual frequency.			
		f_i	B	$(f_i-\hat{f}_i)$
				$(f_i-\hat{f}_i)^2/f_i$
	Note: do not use values of $\hat{f}_i<5$ for step 4; rather, pool adjacent classes to obtain values of $\hat{f}_i \geq 5$; do not forget to also pool the corresponding f_i values (see statistics textbook).			

Program Listing

			COMMENTS
001	001	*LBLo 21 16 11	
	002	P≓S 16-51	
	003	CLRG 16-53	
	004	STO0 35 00	
	005	P≓S 16-51	
	006	CLRG 16-53	
	007	CLX -51	
	008	RTN 24	
	009	*LBLC 21 13	
010	010	P≓S 16-51	
	011	STO1 35 01	
	012	P≓S 16-51	
	013	R↓ -31	
	014	STOA 35 11	
	015	R↓ -31	
	016	STO1 35 01	
	017	1 01	
	018	RTN 24	
	019	*LBLD 21 14	
020	020	P≓S 16-51	
	021	STO2 35 02	
	022	P≓S 16-51	
	023	R↓ -31	
	024	STOB 35 12	
	025	R↓ -31	
	026	STO2 35 02	
	027	2 02	
	028	RTN 24	
	029	*LBLE 21 15	
030	030	P≓S 16-51	
	031	STO3 35 03	
	032	P≓S 16-51	
	033	R↓ -31	
	034	STOC 35 13	
	035	R↓ -31	
	036	STO3 35 03	
	037	3 03	
	038	RTN 24	
	039	*LBLo 21 16 13	
040	040	P≓S 16-51	
	041	STO4 35 04	
	042	P≓S 16-51	
	043	R↓ -31	
	044	STOD 35 14	
	045	R↓ -31	
	046	STO4 35 04	
	047	4 04	
	048	RTN 24	
	049	*LBLd 21 16 14	
050	050	P≓S 16-51	
	051	STO5 35 05	
	052	P≓S 16-51	
	053	R↓ -31	
	054	STOE 35 15	
	055	R↓ -31	
	056	STO5 35 05	
	057	5 05	
	058	RTN 24	
	059	*LBLe 21 16 15	
	060	P≓S 16-51	
060	061	STO6 35 06	
	062	P≓S 16-51	
	063	R↓ -31	
	064	STOI 35 46	
	065	R↓ -31	
	066	STO6 35 06	
	067	6 06	
	068	RTN 24	
	069	*LBLA 21 11	
	070	STO7 35 07	
070	071	0 00	
	072	STO0 35 00	
	073	RCL1 36 01	
	074	RCLA 36 11	
	075	P≓S 16-51	
	076	RCL1 36 01	
	077	P≓S 16-51	
	078	GSB1 23 01	
	079	GSB3 23 03	
	080	RCL2 36 02	
080	081	RCLB 36 12	
	082	P≓S 16-51	
	083	RCL2 36 02	
	084	P≓S 16-51	
	085	GSB1 23 01	
	086	GSB2 23 02	
	087	GSB3 23 03	
	088	RCL3 36 03	
	089	RCLC 36 13	
	090	P≓S 16-51	
090	091	RCL3 36 03	
	092	P≓S 16-51	
	093	GSB1 23 01	
	094	GSB2 23 02	
	095	GSB3 23 03	
	096	RCL4 36 04	
	097	RCLD 36 14	
	098	P≓S 16-51	
	099	RCL4 36 04	
	100	P≓S 16-51	
100	101	GSB1 23 01	
	102	GSB2 23 02	
	103	GSB3 23 03	
	104	RCL5 36 05	
	105	RCLE 36 15	
	106	P≓S 16-51	
	107	RCL5 36 05	
	108	P≓S 16-51	
	109	GSB1 23 01	
110	110	GSB2 23 02	
	111	GSB3 23 03	
	112	RCL6 36 06	

COMMENTS

REGISTERS

0 $\Sigma \hat{f}_i$	1 μ_1	2 μ_2	3 μ_3	4 μ_4	5 μ_5	6 μ_6	7 x_i	8 δ_i	9 μ_i
S0 Δx	S1 N_1	S2 N_2	S3 N_3	S4 N_4	S5 N_5	S6 N_6	S7	S8 N_i	S9
A δ_1	B δ_2	C δ_3	D δ_4	E δ_5	I δ_6				

Program Listing

				COMMENTS
	113	RCLI	36 46	
	114	P≓S	16-51	
	115	RCL6	36 06	
	116	P≓S	16-51	
	117	GSB1	23 01	
	118	GSB2	23 02	
	119	GSB3	23 03	
120	120	GTO4	22 04	
	121	*LBL1	21 01	
	122	P≓S	16-51	
	123	STO8	35 08	
	124	P≓S	16-51	
	125	R↓	-31	
	126	STO8	35 08	
	127	R↓	-31	
	128	STO9	35 09	
	129	RTN	24	
130	130	*LBL2	21 02	
	131	RCL8	36 08	
	132	X=0?	16-43	
	133	GTO4	22 04	
	134	RTN	24	
	135	*LBL3	21 03	
	136	RCL7	36 07	
	137	RCL9	36 09	
	138	-	-45	
	139	RCL8	36 08	
140	140	÷	-24	
	141	X^2	53	
	142	2	02	
	143	÷	-24	
	144	CHS	-22	
	145	e^x	33	
	146	RCL8	36 08	
	147	÷	-24	
	148	Pi	16-24	
	149	2	02	
150	150	x	-35	
	151	$\sqrt{X}$	54	
	152	÷	-24	
	153	P≓S	16-51	
	154	RCL8	36 08	
	155	x	-35	
	156	RCL0	36 00	
	157	x	-35	
	158	P≓S	16-51	
	159	ST+0	35-55 00	
160	160	RTN	24	
	161	*LBL4	21 04	
	162	RCL0	36 00	
	163	SPC	16-11	
	164	PRTX	-14	
	165	R/S	51	
	166	*LBLB	21 12	
	167	RCL0	36 00	
	168	-	-45	
	169	PRTX	-14	
170	170	X^2	53	
	171	RCL0	36 00	
	172	÷	-24	
	173	PRTX	-14	
	174	RTN	24	
	175	R/S	51	
180				
190				
200				
210				
220				

COMMENTS

LABELS

A $X_i \rightarrow f_i$	B	C $\mu_1, \delta_1, N_1^{\rightarrow}$	D $\mu_2, \delta_2, N_2^{\rightarrow}$	E $\mu_3, \delta_3, N_3^{\rightarrow}$	0
a initialize	b	c $\mu_4, \delta_4, N_4^{\rightarrow}$	d $\mu_5, \delta_5, N_5^{\rightarrow}$	e $\mu_6, \delta_6, N_6^{\rightarrow}$	1
0	1 transfer μ, δ, N	2 check no. components	3 computed partial f_i	4 output f_i	2
5	6	7	8	9	3

FLAGS

SET STATUS

FLAGS	ON	OFF
0	☐	☒
1	☐	☒
2	☐	☒
3	☐	☒

TRIG	
DEG	☒
GRAD	☐
RAD	☐

DISP	
FIX	☒
SCI	☐
ENG	☐
n	3